새로운 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘

류종훈[†] · 김대경[††] · 박장식[†††] · 손경식[††††]

요 약

본 논문에서는 확장 스펙트럼 차감 알고리즘으로 처리된 신호의 추정 신호 대 잡음비를 이용한 새로운 음성 활동 검출법을 제안한다. 확장 스펙트럼 차감 알고리즘의 Wiener 필터 출력 신호에서 신호 대 잡음비를 추정하기 위한 Wiener 필터를 하나 더 둠으로써 음성 활동을 검출한다. 제안하는 음성 활동 검출기는 계산량이 많지 않으며 낮은 신호 대 잡음비에서도 잘 동작했다. 제안하는 음성 활동 검출기의 응용으로 Boll의 스펙트럼 차감 알고리즘에 제안하는 음성 활동 검출기를 적용한 다음 확장 스펙트럼 차감 알고리즘과 비교하였다. 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘은 음성/비음성 구간 모두에서 확장 스펙트럼 차감 알고리즘보다 우수한 성능을 보였다.

Boll's Spectral Subtraction Algorithm by New Voice Activity Detection

Ryu Jong Hoon[†], Kim Dae Kyung[††], Park Jang Sik[†††] and Son Kyung Sik[††††]

ABSTRACT

In this paper, a new voice activity detection method estimating SNR of enhanced speech with extended spectral subtraction (ESS) is proposed. Voice activity detection is performed by putting an second Wiener filter behind an Wiener filter used in the ESS to estimate speech and noise power of output signal of first Wiener filter. The proposed voice activity detection method does not require many computational loads and performs well under severe input SNR. Boll's spectral subtraction algorithm with proposed voice activity detection was compared to ESS under several noise environment having different time-frequency distributions. During speech and non-speech activity, performance of Boll's spectral subtraction algorithm with proposed voice activity detection is superior to that of ESS.

1. 서 론

사용자의 편리함과 더불어 통신시스템 및 음성인식 시스템 등의 많은 응용 분야에서 전처리를 위하여 원하는 음성 신호가 배경 잡음과 섞여 있을 때 잡음을 제거함으로써 음질을 개선하는 것이 필요하다. 배경 잡음이 깨끗한 음성에 대하여 부가적이며 상관관계가 없다고 가정할 수 있을 때 잡음 제거를 위한 방법으로 적응 잡음 제거법과 스펙트럼 차감법 등이 있다.

1970년대 중반 이후 다양한 배경 잡음 환경에서 적응 잡음 제거기(adaptive noise canceller : ANC)를 사용함으로써 상당한 신호 대 잡음비 개선 효과를 보여주는 결과들이 제시되었다. 그러나, 적응 잡음 제거기의 전형적인 구조로 인하여 기준 입력단에 입력되는 주입력 신호의 누화 성분을 크게 감쇠시키면서 주입력 신호와 기준 입력 신호 사이에서의 상관관계를 크게 하는 것은 매우 어렵기 때문에 실제 적용에는 많은 제약이 있다[1,2].

[†] LG 전자 가산 사업장 미디어 기기실
[††] 동의공업대학 영상정보과 조교수
[†††] 정회원, 동의공업대학 영상정보과 조교수
[††††] 부산대학교 전자공학과 교수

잡음 섞인 음성이 단일 채널을 통하여 입력될 경우 스펙트럼 차감법이 널리 쓰인다. Boll[3]은 음성 활동 검출기를 통하여 음성/비음성 구간을 구분하고 음성 스펙트럼과 잡음 스펙트럼을 추정한 후, 스펙트럼 크기 평균(magnitude averaging), 반파 정류(half wave rectification), 잔여 잡음 제거를 통해 스펙트럼 추정 오차를 줄인 다음, 비음성 구간에서 추정된 잡음 스펙트럼을 음성 구간에서 차감하는 스펙트럼 차감 알고리즘을 제안했다. 그러나 실제 환경에서 음성/비음성 구간 구분의 어려움으로 인해 음성 활동 검출기를 사용하는 Boll의 스펙트럼 차감 알고리즘의 성능이 저하된다.

Ephraim과 Malah[4]는 MMSE(minimum mean square estimator)를 스펙트럼 차감 알고리즘에 적용함으로써 잔여 음악성 잡음(residual musical noise)이 발생하지 않게 하였다. 그러나 MMSE를 사용함으로 인해 알고리즘이 복잡해지고 계산량이 많아지며, 또한 신호 대 잡음비가 낮은 신호에서는 성능이 좋지 못하다는 단점이 있다.

Martin[5]은 잡음 섞인 음성의 짧은 구간에서 음성의 특성 (부밴드의 계곡값)을 이용하여 음성 활동 검출기를 사용하지 않고도 잡음을 제거할 수 있음을 보였다. 이 알고리즘은 잡음이 섞인 음성의 짧은 구간 전력 추정값의 계곡들이 배경 잡음의 긴 구간의 추정값으로 사용되어질 수 있다는 가정에 기초를 두고 있다. 그러나 이 가정이 위배되는 환경에서는 알고리즘의 성능은 떨어지게 된다.

Sovka[6,7]는 Wiener 필터를 이용하여 음성/비음성 구간의 구분 없이 잡음을 추정하여 제거하였다. Sovka가 제안한 확장 스펙트럼 차감 알고리즘은 Wiener 필터를 이용함으로써 음성 활동 검출기를 필요로 하지 않으며 계산량이 적다는 장점이 있지만 잔여 음악성 잡음을 처리하지 않는 문제가 있다.

본 논문에서는 확장 스펙트럼 차감 알고리즘의 Wiener 필터를 통과한 신호에서 신호 대 잡음비를 추정하여 음성 활동 검출을 함으로써 계산량이 적은 음성 활동 검출법을 제안하였다. 확장 스펙트럼 차감 알고리즘의 Wiener 필터를 통과한 신호는 잡음이 상당 부분 제거되었으나 잔여 음악성 잡음이 있으므로 잡음 섞인 음성으로 볼 수 있다. 확장 스펙트럼 차감 알고리즘에서 사용된 Wiener 필터 출력단에 Wiener 필터를 한번 더 사용하여 Wiener 필터를 통과한 신호의 음성 전력과 잡음 전력을 추정한다. 추정된 음성 전력과 잡음 전력으로 추정 신호 대 잡음비를 구한 다음 실험적으로 구해진 문턱값보다 클 경우 음성 구간, 작을 경우 비음성 구간으로 판단한다. 제안하는 음성 활동 검출법의 응용으로 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘을 시뮬레이션하고 확장 스펙트럼 차감 알고리즘의 시뮬레이션 결과와 비교하였다.

2. 스펙트럼 차감 알고리즘

시간 영역의 신호를 주파수 영역으로 변환하여 각 프레임의 입력 신호 스펙트럼 크기와 잡음 스펙트럼 크기를 추정하고, 추정된 입력 신호 스펙트럼 크기에서 잡음 스펙트럼 크기를 차감하여 음성 스펙트럼 크기를 추정한다. 이 때 음성 신호가 짧은 구간동안 정적이라는 가정하에서 표본화 주파수가 8 kHz일때 프레임 길이는 보통 160 샘플에서 256 샘플사이의 값이 된다. 스펙트럼 추정 오차를 줄이기 위해 입력 신호는 1/2 중첩되는 동일한 길이를 가진 프레임으로 나누어지고, Hanning 이나 Hamming 윈도우 등이 가중치 함수로써 이용되어진다[8].

음성 신호의 위상은 사람의 청각에 큰 영향을 주지 않는다는 가정하에서 추정된 음성 스펙트럼 크기를 잡음 섞인 음성 신호의 위상과 결합한 다음 시간 영역으로 변환하면 시간 영역에서의 음성 신호 를 얻을 수 있다. 깨끗한 음성과 배경 잡음은 부가적이고 상관 관계가 없으며 짧은 시간동안 정적(locally stationary)이라고 가정하면, 식 1과 같이 표현할 수 있다.

$$X(\omega;m) = S(\omega;m) + N(\omega;m) \quad (1)$$

식 1에서 $X(\omega;m)$, $S(\omega;m)$, $N(\omega;m)$은 각각 잡음 섞인 음성, 깨끗한 음성, 잡음의 단구간 푸리에변환(short-time discrete time Fourier transform : st-DTFT)이며, m은 프레임 인덱스이다. 그리고, 추정 음성 스펙트럼의 위상을 입력 신호 스펙트럼의 위상과 같게 하면 식 2와 같이 표현된다.

$$\hat{S}(\omega;m) = [|X(\omega;m)| - |\hat{N}(\omega;m)|]e^{j\varphi_x(\omega;m)} \quad (2)$$

음성 활동 검출기는 많은 음질 개선 알고리즘에서 매우 중요하다 음성/비음성 구간의 판단 기준으로는 신호 대 잡음비, 영교차율(zero crossing rate), LPC

거리차 등이 있는데 입력 신호 대 잡음비가 낮을 때에는 이러한 방법들을 사용하더라도 신뢰성이 다소 떨어진다. 이점이 음성 활동 검출기를 이용하는 알고리즘에서 성능을 저하시키는 심각한 원인이 된다[9,10]. 낮은 신호 대 잡음비 환경에서도 우수한 성능을 보이는 음성 활동 검출기도 있지만, 알고리즘의 복잡함과 많은 계산량으로 인하여 실시간 환경에서 동작하기에는 어려움이 있다[9-13].

음성활동 검출기의 성능 평가는 응용에 따라 달라지게 된다. 음성 부호화기의 경우, 비음성 구간을 음성으로 판단하는 것은 음성 부호화기의 성능을 크게 저하시키지 않지만, 음성 구간을 비음성구간으로 판단하는 것은 음성을 왜곡시키므로, 음성 부호화기의 성능을 크게 저하시킨다. 그러나, 적응 반향제거기에서는 원단 화자의 음성 구간을 비음성 구간으로 판단할 경우 반향을 제거하지 못하므로 성능이 약간 저하되지만, 비음성 구간을 음성으로 판단할 경우 적응 필터가 오동작을 일으키게 되므로 성능이 크게 저하된다. 음성 활동 검출기를 사용하는 스펙트럼 차감 알고리즘의 경우 비음성 구간을 음성 구간으로 판단하는 경우, 정확한 잡음 추정이 이루어지지 않으므로 전체 알고리즘의 성능이 저하되고, 음성 구간을 비음성 구간으로 판단하는 경우, 음성 왜곡이 있게 되어 전체 알고리즘의 성능이 저하된다.

2.1 Boll 의 스펙트럼 차감 알고리즘

Boll[3]은 입력 신호 스펙트럼으로부터 잡음 스펙트럼을 차감함으로써 잡음 제거가 가능하다는 가정하에 추정 잡음 스펙트럼 $|\hat{N}(\omega;m)|$을 비음성 구간에서 얻어진 평균 잡음 스펙트럼 $\mu(\omega;m)$로 대체하였다. Boll이 제안한 방법은 그림 1과 같다.
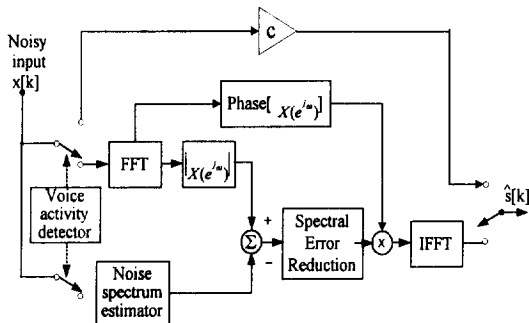
그림 1. Boll의 스펙트럼 차감 알고리즘

$$\hat{S}(\omega;m) = [|X(\omega;m)| - \mu(\omega;m)]e^{j\varphi_x(\omega;m)} \quad (3)$$

여기서 $\mu(e^{j\omega})$은 식 4와 같다.

$$\mu(\omega;m) = E\{|N(\omega;m)|\} \quad (4)$$

음성 스펙트럼 추정 오차를 감소시키기 위하여 스펙트럼 크기 평균(magnitude averaging), 반파 정류, 잔여 잡음 제거의 세 가지 방법을 사용하였으며 식 5와 같이 T를 구하여 주어진 프레임에서 의 음성 활동 검출 판정 기준으로 이용하였다.

$$T = 20\log_{10}[\ \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\hat{S}(\omega;m)}{\mu(\omega;m)}\right| d\omega\ ] \quad (5)$$

이 때, T가 실험적으로 구해진 문턱값 12 dB보다 크면, 주어진 프레임에서 음성 활동이 있는 것으로 판정하여 아무런 처리를 하지 않고, T가 12 dB보다 작으면 음성 활동이 없는 것으로 판정하여 이 프레임에서는 추정 음성 스펙트럼을 감쇠된 입력 신호 스펙트럼으로 대체하였다.

$$\hat{S}(\omega;m) = \begin{cases} \hat{S}(\omega;m)\ , & T \geq -12\,dB \\ cX(\omega;m)\ , & T < -12\,dB \end{cases} \quad (6)$$

여기서, 식 6의 c는 $20\log_{10} c$ = -30 dB를 만족하는 값이고, 약 0.03162이다.

2.2 확장 스펙트럼 차감 알고리즘

Sovka[6,7]는 스펙트럼 차감 알고리즘과 반복 Wiener 필터링을 결합한 확장 스펙트럼 차감 알고리즘을 제안하였다. 확장 스펙트럼 차감 알고리즘의 주요한 특성은 음성 활동 검출기를 필요로 하지 않는다는 것과 음성 구간에서도 배경 잡음 추정을 지속적으로 갱신할 수 있다는 것이다.

Wiener 필터를 이용한 스펙트럼 차감 알고리즘은 입력 신호를 FFT한 후 Winer 필터를 통과시켜서 얻어진 값들을 IFFT하여 음질이 개선된 신호를 얻어낸다. 여기서 Winer 필터의 주파수 특성은 식 7과 같다.

$$H^2(e^{j\omega}) = \frac{|X(e^{j\omega})|^2 - |\bar{N}(e^{j\omega})|^2}{|X(e^{j\omega})|^2} \quad (7)$$

여기서 $|\bar{N}(e^{j\omega})|^2$는 비음성 구간에서 구해진 평균 잡음 전력 스펙트럼이다.

일반적인 Wiener 필터에 기반을 둔 스펙트럼 차

감 알고리즘과 확장 스펙트럼 차감 알고리즘은 몇 가지 중요한 차이점이 있다. 첫째, 일반적인 Wiener 필터는 입력 신호에서 음성 스펙트럼을 추정하는 반면 확장 스펙트럼 차감 알고리즘에서 이용되는 Wiener 필터는 입력 신호에서 잡음 스펙트럼을 추정한다. 둘째, 입력 신호가 아닌 출력 신호로 Wiener 필터를 근사시킨다. 그러므로 음성 활동 검출기가 필요하지 않다. 세번째로, 확장 스펙트럼 차감 알고리즘 구조에서는 되먹임(feed-back)이 이루어지므로, 확장 스펙트럼 차감 알고리즘은 반복 Wiener 필터링으로 볼 수 있다. 확장 스펙트럼 차감 알고리즘의 구조는 그림 2와 같다.
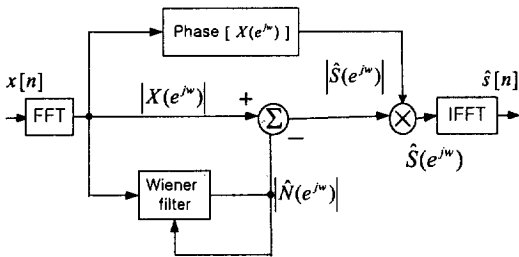
그림 2. 확장 스펙트럼 차감 알고리즘

확장 스펙트럼 차감 알고리즘에서 Wiener 필터의 주파수 응답은 식 8로 나타낼 수 있다.

$$H_n(e^{j\omega}) = \left(\frac{|\overline{N}_{n-1}(e^{j\omega})|^2}{|\overline{N}_{n-1}(e^{j\omega})|^2 - |\overline{S}_{n-1}(e^{j\omega})|^2} \right)^{1/2} \quad (8)$$

입력 신호에서 Wiener 필터를 통해 구해진 추정 잡음 스펙트럼은, 다음 식에 의해 지수적으로 평균된다.

$$|\overline{N}_{n+1}(e^{i\omega})| = p|\overline{N}_n(e^{i\omega})| + (1-p)|\hat{N}_n(e^{i\omega})| \quad (9)$$

$$|\overline{S}_n(e^{i\omega})| = |X_n(e^{i\omega})| - |\overline{N}_n(e^{i\omega})| \quad (10)$$

식 10에서 p의 값은 전체 알고리즘의 동작에 중요한 영향을 끼친다. 만약, $|\overline{N}(e^{jw})|$과 $|\hat{N}(e^{jw})|$에서 상대적으로 완만한 변화가 나타나고, 이보다 빠른 변화는 $|\overline{S}(e^{jw})|$에서 나타난다고 가정한다면, p의 값을 적절하게 정할 수 있고 결과적으로 전체 알고리즘이 잘 동작하게 된다.

3. 제안하는 음성 활동 검출법 (VAD)

Boll의 알고리즘은 비음성 구간에서의 잔여 잡음 제거를 통해 비음성 구간에서 잔여 음악성 잡음을 감소시키는 장점이 있는 반면, 계산량이 적으며 낮은 신호 대 잡음비에서도 신뢰성 있는 음성 활동 검출기를 필요로 하게 된다. 반면, 확장 스펙트럼 차감 알고리즘은 계산량이 적은 장점이 있으나 비음성 구간에서의 잔여 음악성 잡음을 처리하지 않는 문제점이 있다. 따라서 확장 스펙트럼 차감 알고리즘의 Wiener 필터를 개선하여 계산량이 적으면서도 낮은 신호 대 잡음비에서도 동작하는 음성 활동 검출기를 제안한다.

확장 스펙트럼 차감 알고리즘에서 입력 신호가 Wiener 필터를 통과하면 잡음 성분이 제거된 신호가 출력된다. 만일 Wiener 필터가 잡음 스펙트럼을 정확하게 추정한다면 추정 음성 스펙트럼 $|\hat{S}_n(e^{jw})|$은 음성 스펙트럼 $|S_n(e^{jw})|$과 같지만, 실제로는 추정 오차가 있으므로 Wiener 필터를 통과한 신호에는 잔여 잡음이 여전히 남아 있다.

추정된 음성 스펙트럼에서 신호 대 잡음비를 구하여 실험적으로 구한 문턱값보다 크면 음성 구간, 작으면 비음성 구간으로 판단한다. 즉, Wiener 필터를 통과한 신호는 잡음 섞인 음성 스펙트럼이므로, Wiener 필터를 하나 더 써서 평균 잡음 스펙트럼과 음성 스펙트럼을 추정한다. 이때, Wiener 필터의 주파수 응답은 아래 식11과 같이 나타낼 수 있다.

$$H'_n(e^{j\omega}) = \left(\frac{|\overline{N}'_{n-1}(e^{j\omega})|^2}{|\overline{N}'_{n-1}(e^{j\omega})|^2 - |\overline{S}'_{n-1}(e^{j\omega})|^2} \right)^{1/2} \quad (11)$$

이때, 추정 평균 음성 스펙트럼과 추정 평균 잡음 스펙트럼은 12, 13 식과 같이 주어진다.

$$|\overline{N}'_{n-1}(e^{i\omega})| = p|\overline{N}'_{n-1}(e^{i\omega})| + (1-p)|\hat{N}'_{n-1}(e^{i\omega})| \quad (12)$$

$$|\overline{S}'_{n-1}(e^{i\omega})| = |\hat{S}_n(e^{i\omega})| - |\overline{N}'_{n-1}(e^{i\omega})| \quad (13)$$

추정된 평균 잡음 스펙트럼과 음성 스펙트럼을 이용하여 추정 프레임 평균 잡음 전력과 추정 프레임 음성 전력을 구하고 이에 의해 추정 출력 신호 대 잡음비를 구한 다음 문턱값과 비교한다.

$$\text{Speech}: \text{if } \left| \frac{\hat{S}'_n(\omega;m)}{\overline{N}'_n(\omega;m)} \right| > \textit{Threshold},$$
$$\text{Non-speech}: \text{else} \quad (14)$$

일반적으로 음성구간에서 신호대 잡음비가 크고, 비음성 구간에서는 신호대 잡음비가 작게 나타나므로 주어진 잡음 조건에 따라 문턱값을 실험적으로

정한다. 이 때, 문턱값이 너무 낮을 경우 비음성 구간에서 음성 검출 성능이 저하되고 문턱값이 너무 높을 경우, 음성 구간에서 음성 검출 성능이 저하된다. 제안하는 스펙트럼 차감법의 구조는 그림 3과 같다.
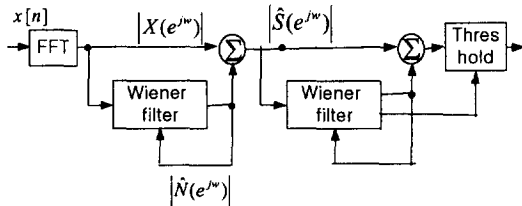
그림 3. 제안하는 음성 활동 검출기를 이용한 스펙트럼 차감법

제안하는 음성 활동 검출기를 스펙트럼 차감 알고리즘에 적용할 경우, 음성 활동 검출기를 필요로 하는 스펙트럼 차감 알고리즘에서는 알고리즘 내부에서 고속 푸리에 변환(FFT)을 하므로 별도로 FFT를 할 필요가 없다.

4. 제안한 음성 활동 검출법을 이용한 시뮬레이션 결과 및 검토

4.1 시뮬레이션 환경

깨끗한 음성에 자동차 잡음, 백색 가우시안 잡음 및 헬리콥터 잡음을 섞은 신호를 입력 신호로 하여 제안하는 음성 활동 검출법에 의한 Boll의 알고리즘과 확장 스펙트럼 차감 알고리즘을 시뮬레이션하였다. 시뮬레이션에서 사용된 깨끗한 음성은 8 kHz로 표본화되고 16비트로 양자화된 저 거기 회계과 부탁해요 라는 ETRI 전화용 데이터이다.

자동차 잡음은 고속도로에서 100km/h로 주행하는 소형승용차의 운전석에서 녹음한 것을 8 kHz로 표본화, 16 bit로 양자화한 데이터이다. 자동차 잡음 녹음시 자동차의 창문을 모두 닫아서 바람에 의한 잡음을 최대한 감소시켰고 주변에 다른 자동차가 지나가지 않을 때 녹음하였으므로 다른 자동차에 의한 잡음의 영향이 거의 없었다. 자동차 잡음은 차량 운행 환경, 차종 등에 따라 다르지만 일반적으로 입력 신호 대 잡음비가 10 dB 이상인 것으로 알려져 있다[14]. 헬리콥터 잡음은 실제 UH-60 헬리콥터 의 조종실에서 녹음한 데이터를 이용하였다. 일반적으로 비행 중의 헬리콥터 잡음은 잡음 지수가 94~98 dB SPL(sound pressure level)범위에 있어, 조종사와 조종사간은 물론 탑승자들 상호간에도 헤드셋이나, 조종사 헬멧에 부착되어 있는 잡음 제거 마이크를 사용하지 않으면 의사 소통이 어려운 상황이다. 또한, 비행 중 반드시 필요한 조종사와 지상 관제사간의 무전 교신을 할 때에 헬리콥터의 잡음이 많이 포함되어 있어 조종사가 비행 중 무선 교신에 집중을 해야 하므로 조종사에게 큰 부담이 된다. 헬리콥터 내부에는 각종 무전기와 항법 장비 등 많은 전자 장비들이 있지만 소음을 제거할 수 있는 장비로는 잡음 제거 마이크만 있다. 헬멧에 부착되어 있는 잡음 제거 마이크의 성능은 평균 감쇠 비율 24 dB NNR(noise reduction rate)이고, 조종사의 잡음 제거 마이크를 통해 출력된 음성 신호의 신호 대 잡음비는 5~10 dB 이다. 시뮬레이션에 사용될 데이터를 얻기 위해서 조종사가 실제 환경에서 사용하는 헬멧에 부착된 잡음 제거 마이크를 이용하였다. 헬리콥터의 고도는 3000 ft(약 915 m), 속도는 순항 속도인 140 kts(약 260 km/h), 온도는 섭씨15도였다. 표본화 주파수가 8 kHz일 때, 일반적으로 윈도우의 길이는 160~256 샘플이 된다. Boll의 스펙트럼 차감 알고리즘과 확장 스펙트럼 차감 알고리즘에서 모두 길이가 256인 윈도우를 1/2만큼 중첩시켰고 가중 윈도우로는 Hanning 윈도우를 사용하였다. 제안하는 음성 활동 검출법에서 음성, 비음성 구간을 구분하기 위한 문턱 신호 대 잡음비는 6으로 두었고 추정 신호 대 잡음비는 각 프레임마다 망각 지수 0.90을 두어 갱신하였다. Boll은 식 6에서 c를 약 0.03162로 두었으나, 본 시뮬레이션에서는 c를 0.1로 두었다. 이 때, c값은 여러 가지로 바꾸어가며 Boll의 스펙트럼 차감 알고리즘의 출력 신호를 청취한 결과 가장 음질이 좋다고 판단되는 경우의 값으로 정하였다. c를 실험적으로 정한 값 이외의 다른 값으로 하였을 경우에 음성 활동 검출기가 정확하게 동작하지 않아서 처리된 음성에 잔여 잡음이 많이 남았다. Boll의 스펙트럼 차감 알고리즘, 확장 스펙트럼 차감 알고리즘에서 잡음 스펙트럼 추정시 망각 지수를 0.93으로 두었다[15].

4.2 성능 평가 기준

본 논문에서는 성능 평가 기준으로 Sovka가 제안한 SSNRE를 사용하였다. Sovka가 제안한 방법은 세그멘털 신호 대 잡음비의 차의 평균을 이용하여

음성의 향상 정도를 구분하는 것이다. 즉 입력 신호의 각 프레임에서 세그멘털 신호 대 잡음비를 구하고 다시 출력 신호의 각 프레임에서 세그멘털 신호 대 잡음비를 구한 후 식 17와 같이 두 값의 차를 구한다.

$$SNR[m] = 10\log\frac{P_S[m]}{P_N[m]} = 10\log\frac{P_S[m]}{P_{X-S}[m]} = 10\log\frac{P_{X-N}[m]}{P_N[m]} \quad (15)$$

$$SSNR = \frac{1}{L}\sum_{m=1}^{L} SNR[m]\text{ , } L\text{ :total frame number} \quad (16)$$

$$SSNRE = SSNR_{out} - SSNR_{in} \quad (17)$$

식 15, 16에서 $P_S[m]$, $P_N[m]$, $SNR[m]$ 은 각각 m번째 프레임에서의 신호 전력, 잡음 전력, 신호 대 잡음비를 나타내고, 식 17에서 $SSNR_{out}$ 은 출력 신호의 세그멘털 신호 대 잡음비, $SSNR_{in}$ 은 입력 신호의 세그멘털 신호 대 잡음비를 나타낸다. SSNRE는 입력과 출력 세그멘털 신호 대 잡음비의 차의 평균으로 값이 클수록 성능이 우수함을 나타낸다.

4.3 제안한 음성 활동 검출법을 이용한 스펙트럼 차감 알고리즘과 기존 알고리즘의 성능 비교

그림 4, 5, 6은 자동차 잡음, 백색 가우시안 잡음, 헬리콥터 잡음 각각의 배경 잡음에 대해서 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘과 확장 스펙트럼 차감 알고리즘을 시뮬레이션 한 것이다. 각 그림에서 (a)는 깨끗한 음성, (b)는 배경 잡음이 섞인 음성, (c)는 확장 스펙트럼 차감 알고리즘 출력 신호, (d)는 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호이다. 그림 (e)에서 실선은 추정 신호 대 잡음비이고, 점선은 문턱 신호 대 잡음비이다. 그림에서 보이듯이 실선이 점선 위에 있을 때 음성 구간으로 판단하였음을 알 수 있다.

배경 잡음이 자동차 잡음인 경우 그림 4에서 알 수 있듯이 비음성 구간에서는 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘과 확장 스펙트럼 차감 알고리즘의 성능이 비슷함을 알 수 있다. 이것은 자동차 잡음의 전력은 대부분 저주파인 50~1200 Hz 사이에 있고, 따라서 확장 스펙트럼 차감 알고리즘의 Wiener 필터에서 추정된 잡음 스펙트럼이 실제 잡음 스펙트럼과 큰 차이가 없기
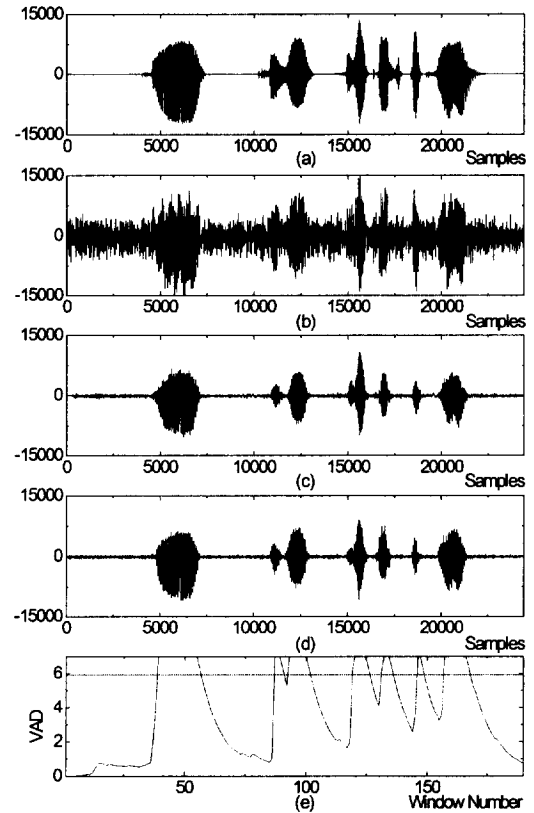
그림 4. 자동차 잡음에 대한 각 알고리즘의 출력 신호 (0dB) (a) 깨끗한 음성 (b) 입력 신호 (c) ESS (d) 제안하는 VAD에 의한 Boll의 알고리즘 (e) VAD

때문이다. 반면, 배경 잡음이 각각 백색 가우시안 잡음과 헬리콥터 잡음인 경우 비음성 구간에서 확장 스펙트럼 차감 알고리즘 출력 신호에 잔여 잡음이 더 많음을 알 수 있다. 이것은 백색 가우시안 잡음과 헬리콥터 잡음에는 고주파 성분이 많고 따라서 확장 스펙트럼 차감 알고리즘의 Wiener 필터에서 추정된 잡음 스펙트럼과 실제 잡음 스펙트럼의 차가 크기 때문이다.

배경 잡음이 자동차 잡음, 백색 가우시안 잡음, 헬리콥터 잡음인 경우 모두에서, 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호가 확장 스펙트럼 차감 알고리즘 출력 신호보다 음성 구간에서 깨끗한 음성파형에 더 유사하다. 그러나 제안하는 음성 활동 검출기가 완벽하게 동작하지 않으므로 제안하는 음성 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호는 음성 전후에서 약간 왜곡된다. 제안하는 음성 활동 검출법에 의한 Boll의

스펙트럼 차감 알고리즘 출력 신호와 확장 스펙트럼 차감 알고리즘 출력 신호를 청취한 결과, 음성 구간의 경우 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호에서 잔여 음악성 잡음이 약간 더 크게 들렸다. 확장 스펙트럼 차감 알고리즘 출력 신호의 경우, 음성/비음성 구간 모두에 잔여 음악성 잡음이 있으므로 비음성 구간에서 음성 구간으로 갈 때, 청취에 크게 거슬리지 않는다. 반면, Boll의 스펙트럼 차감 알고리즘 출력 신호는 비음성 구간의 잔여 음악성 잡음이 감쇠된 배경 잡음으로 대체되었으므로, 배경 잡음만 있던 비음성 구간에서 잔여 음악성 잡음이 있는 음성 구간으로 갈 때, 잔여 음악성 잡음이 부각되어 들리게 된다.

제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호에서 배경 잡음이 백색 가우시안 잡음일 때 음성 구간에서 잔여 음악성 잡음이 가장 크게 부각되어 들렸고 배경 잡음이 자동차 잡음일 때 가장 작게 들렸다. 이것은 배경 잡음이 자동차 잡음인 경우, 잡음 전력이 대부분 저주파에 집중되어 있고 따라서 음성 구간에서 잔여 음악성 잡음 일부분이 음성에 마스킹되어 청취자에게는 실제보다 작게 들리는 반면, 배경 잡음이 백색 가우시안 잡음인 경우 고주파 성분이 많이 있고, 따라서 음성 구간의 잔여 음악성 잡음이 음성에 의해 마스킹되지 않기 때문이다. 음성/비음성 구간 전체적으로는 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호가 확장 스펙트럼 차감 알고리즘 출력 신호보다 잔여 음악성 잡음이 작았고 음질이 우수했다.

그림 7은 자동차 잡음, 백색 가우시안 잡음, 헬리콥터 잡음의 각 경우에 대하여 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘과 확장 스펙트럼 차감 알고리즘 출력 신호의 SNRE를 그린
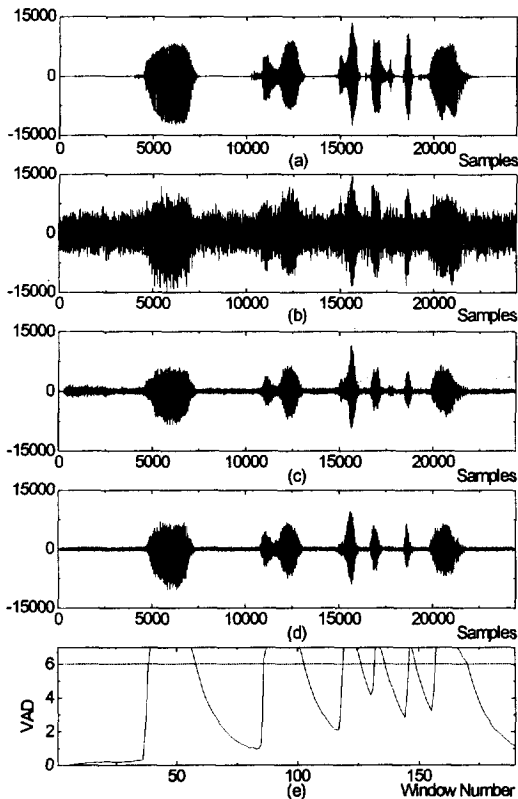
그림 5. 백색 가우시안 잡음에 대한 각 알고리즘의 출력 신호 (0dB)
(a) 깨끗한 음성 (b) 입력 신호 (c) ESS (d) 제안하는 VAD에 의한 Boll의 알고리즘 (e) VAD
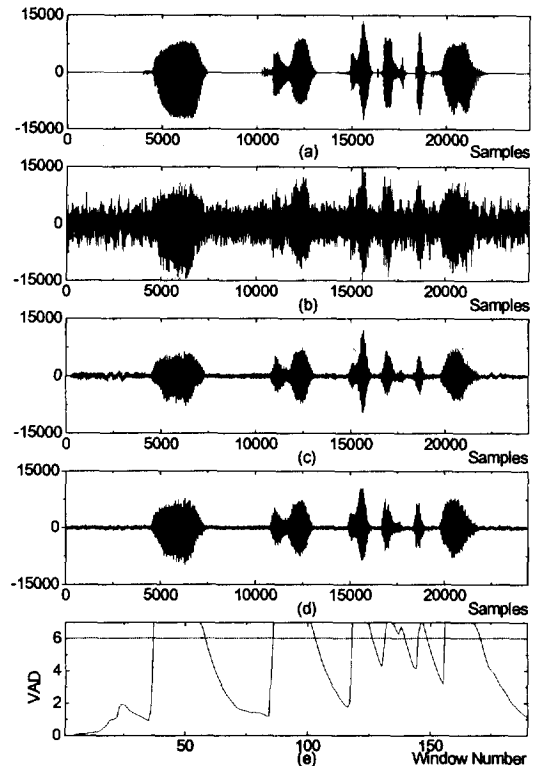
그림 6. 헬리콥터 잡음에 대한 각 알고리즘의 출력 신호 (0dB)
(a) 깨끗한 음성 (b) 입력 신호 (c) ESS (d) 제안하는 VAD에 의한 Boll의 알고리즘 (e) VAD
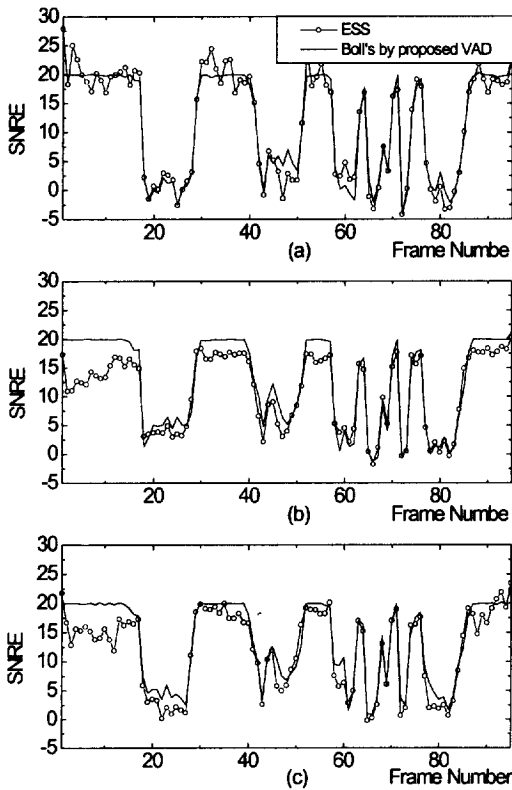
그림 7. 각 알고리즘 출력 신호의 SNRE
(a) 자동차 잡음 (b) 백색 가우시안 잡음 (c) 헬리콥터 잡음

것이다. 배경 잡음이 자동차 잡음인 경우, 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호의 SNRE는 음성/비음성 구간 모두에서 확장 스펙트럼 차감 알고리즘 출력 신호의 SNRE와 비슷하다. 이것은 그림 4에서 살펴본 바와 같이 자동차 잡음은 저주파 대역에 전력이 대부분 집중되어 있어, 확장 스펙트럼 차감 알고리즘의 Wiener 필터가 잡음 스펙트럼을 잘 추정하기 때문이다. 배경 잡음이 백색 가우시안 잡음, 헬리콥터 잡음인 경우, 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호의 SNRE는 비음성 구간에서 확장 스펙트럼 차감 알고리즘 출력 신호의 SNRE보다 약 2~3 dB 크다. 이것은 확장 스펙트럼 차감 알고리즘에서는 잔여 잡음 처리를 하지 않는 반면, Boll의 스펙트럼 차감 알고리즘에서는 후처리를 하여 잔여 음악성 잡음을 감쇠된 배경 잡음으로 대체했기 때문이다. 이 때 음성 구간에서는 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호가 확장 스펙트럼 차감 알고리즘 출력 신호보다 SNRE가 약 1dB 정도 크다.

Boll의 스펙트럼 차감 알고리즘을 시뮬레이션 할 때 식 5의 감쇠 상수 c를 0.1로 했다. 이것은 Boll이 제안한 값인 0.03162로 했을 경우 비음성 구간에서 음성 구간으로 가면서 잔여 음악성 잡음이 부각되어 들리게 되어 오히려 음질이 저하되었기 때문이다.

따라서 음질이 가장 좋을 경우의 c를 실험적으로 구하였다. 만약, 감쇠 상수 c를 Boll이 제시한 대로 0.03162로 했을 경우 전체적인 음질은 저하되나 이 때 출력 신호의 SNRE는 비음성 구간에서 약 30 dB가 되어 c를 0.1로 했을 때의 20 dB보다 약 10 dB 증가한다.

그림 8은 3 dB에서 9 dB까지 3 dB간격으로 입력 신호 대 잡음비를 바꾸어 가며 자동차 잡음, 백색 가우시안 잡음, 헬리콥터 잡음에 대해 제안하는 음성
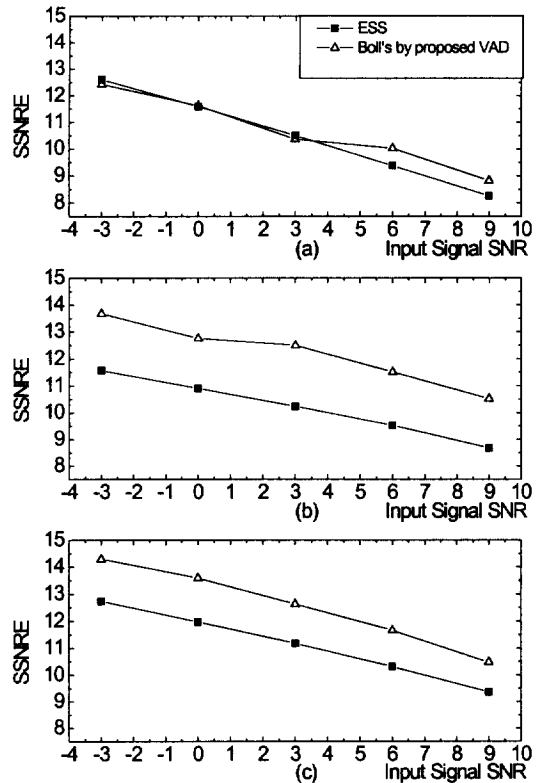
그림 8. 입력 신호 대 잡음비에 따른 각 알고리즘 출력 신호의 SSNRE
(a) 자동차 잡음 (b) 백색 가우시안 잡음 (c) 헬리콥터 잡음

활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호와 확장 스펙트럼 차감 알고리즘 출력 신호의 SSNRE를 그린 것이다. 그림에서 알 수 있듯이 SSNRE의 관점에서 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘이 확장 스펙트럼 차감 알고리즘보다 우수한 성능을 보임을 알 수 있다.

한편, 그림 7에서는 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘과 확장 스펙트럼 차감 알고리즘은 음성 구간에서 비슷한 성능을 보였다. 따라서, 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호가 확장 스펙트럼 차감 알고리즘 출력 신호보다 SSNRE가 큰 것은 대부분 비음성 구간에서의 잔여 잡음 처리에 기인함을 알 수 있다.

5. 결 론

본 논문에서는 확장 스펙트럼 차감 알고리즘의 Wiener 필터를 통과한 신호의 신호 대 잡음비를 추정하여 입력 신호의 음성/비음성 구간을 판단하는 새로운 음성 활동 검출법을 제안하였다. 확장 스펙트럼 차감 알고리즘에서 Wiener 필터를 통과한 신호는 잔여 잡음을 포함하고 있으므로, 잡음 섞인 음성이라 볼 수 있다. 이 신호의 신호 대 잡음비를 추정하고 문턱값과 비교함으로써 Wiener 필터를 통과하기 전 입력 신호의 음성/비음성 검출이 가능해진다. Wiener 필터를 통과한 신호의 음성 스펙트럼 전력과 평균 잡음 스펙트럼 전력을 추정하기 위해 확장 스펙트럼 차감 알고리즘의 Wiener 필터 다음에 Wiener 필터를 하나 더 두었다.

제안하는 음성 활동 검출법의 응용으로, 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘을 시뮬레이션하고 확장 스펙트럼 차감 알고리즘의 시뮬레이션 결과와 비교하였다. 입력 신호 대 잡음비가 0 dB일 때, 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘은 확장 스펙트럼 차감 알고리즘에 비해 음성 구간에서는 SNRE가 약 1 dB 높았고, 비음성 구간에서는 비슷하거나 약 2~3 dB 정도 높았다. 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘과 확장 스펙트럼 차감 알고리즘 출력 신호를 청취하면, 제안하는 음성 활동 검출법에 의한 Boll의 스펙트럼 차감 알고리즘 출력 신호는 음성 구간에서 잔여 음악성 잡음은 약간 증가하였지만 음성 신호전력이 크기 때문에 잡음의 증가 효과가 적게 나타나고, 비음성 구간에서 잔여 음악성 잡음은 많이 감소되어 전체적으로 개선된 음질이 더 우수했다.

참 고 문 헌

[1] W. A. Harrison and J. S. Lim, A new application of adaptive noise cancellation, *IEEE Transaction on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 1, pp. 21-27, Feb. 1986.

[2] M. M. Goulding and J. S. Bird, Speech enhancement for mobile telephony, *IEEE Transactions on Vehicular Technology*, vol.39, no.4, pp. 316-326, Nov. 1990.

[3] S.F.Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, no. 2, pp.113120, Apr. 1979.

[4] Y. Ephraim and D.Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no.6, pp. 443-445, Apr. 1985.

[5] R. Martin, "Spectral subtraction based on minimum statistics," *In Proceedings of EUSIPCO-94, Seventh European Signal Processing Conference*, pp. 1182-1185. 1994.

[6] P. Sovka, P. Pollak, and J. Kybic, "Extended spectral subtraction", *European Conference on Signal Processing and Communication '96*. pp. 454-459, Sept. 1996.

[7] P. Sovka, "Extended spectral subtraction-description and preliminary results," *Research report R95-2*, CTU-Faculty of Electrical Engineering, Prague, 1995.

[8] M. H. Hayes, *Statistical Digital Signal Proc-*

essing and Modelling, John Wiley & Sons, Inc., 1996

[9] K. El-Maleh and P.Kabal, Comparison of voice activity detection algorithms for wireless personal communicatioin systems, *In Proceedings of the Canadian Conference on Electrical and Computer Engineering*, pp. 470-473, May 1997.

[10] L. R. Rabiner and Marvin R. Sambur, Application on an LPC distance measure to the voiced-unvoiced-silence detection problem, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(4), 1977.

[11] L. Karray, C. Mokbel, and J. Monne, Solution for robust speech/non-speech detection in wireless environment, *Proceedings of the 1998 IEEE 4th Workshop Interactive Voice Technology for Telecommunications Applications*, pp. 166-170, 1998.

[12] N. Mesa, J. M. Bilbao. A and L-Solano, An improved speech endpoint detection system in noisy environments by means of third-spectra, *IEEE Signal Processing Letters*, vol 6. no. 9, pp. 224-226, 1999.

[13] D. Wu et. al, A robust speech detection algorithm for speech activated hands-free applications, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol 4. pp. 2407-2410, 1999.

[14] 김대경, 심리 음향 기준을 이용한 음질 개선과 음성 활동 검출, 부산대학교 박사 학위 논문, feb. 2000

[15] 이정화, 프레임 평균과 잔여 잡음 제거를 이용한 확장 주파수 차감 알고리즘의 성능 향상, 부산대학교 석사 학위 논문, Feb. 1999.

류 종 훈

1998년 2월 부산대학교 전자공학과 (공학사)
2000년 2월 부산대학교 전자공학과 (공학석사)
2000년 3월~현재 LG 전자 가산사업장 미디어 기기실
관심분야 : 음성 신호처리, 디지털 신호처리, 음향 신호처리

김 대 경

1992년 2월 부산대학교 전자공학과(공학사)
1995년 2월 부산대학교 전자공학과(공학석사)
2000년 2월 부산대학교 전자공학과(공학박사)
1998년 3월~현재 동의공업대학 영상정보과 조교수
관심분야 : 디지털 신호처리, 음성 신호처리, 멀티미디어 통신

박 장 식

1992년 2월 부산대학교 전자공학과(공학사)
1994년 2월 부산대학교 전자공학과(공학석사)
1999년 2월 부산대학교 전자공학과(공학박사)
1997년 3월~현재 동의공업대학 영상정보과 조교수
관심분야 : 음성 및 음향 신호 처리, 멀티미디어 통신, 입체음향

손 경 식

1973년 2월 부산대학교 전자공학과 졸업(공학사)
1977년 8월 부산대학교 전자공학과(공학석사)
1991년 8월 경북대학교 전자공학과 졸업(공학박사)
1979년~현재 부산대학교 전자공학과 교수
관심분야 : 디지털 신호처리, 적응 신호처리, 음성 및 음향 신호처리